S-181E
C2

Cluster Listing Check Program for the Redesigned LFS Sample

M. Joncas

Census Household Survey Methods Division

September 1985

CLUSTER LISTING CHECK PROGRAM

Introduction

The 1981 Post Censal Redesign of the LFS lead to a complete re-specification of the sample (Singh, Drew, Choudhry, 1985). The new sample was introduced between the months of October 1984 and March 1985, (Mayda, Drew, Lindeyer, 1985) during which period lists of dwellings had to be created in the field and data captured for over 11,000 new clusters. The conditions under which the various operations involved took place (training new staff, familiarization with new types of clusters, etc.) gave rise to a program whose purpose was to check the quality of work done.

Description of the Program

The primary purpose of the program was to measure and evaluate the overall quality of the cluster dwelling lists. In order to measure quality, 2 main types of errors were identified. The first type occurs when a dwelling is erroneously included in the cluster listing. Conversely, the second type of error occurs when a dwelling is omitted from the cluster listing. In an attempt to better identify the nature of these errors, the 2 types of errors have been broken down into 4 and 6 categories, respectively. These categories are as follows:

Dwellings Erroneously Included

(3.1) - misinterpretation of cluster boundaries
(3.2) - dwellings converted into commercial facilities
(3.3) - dwellings demolished or destroyed by fire
(3.4) - dwellings erroneously included

Dwellings Inadvertently Omitted

(1.1) - misinterpretation of cluster boundaries
(1.2) - section of street omitted
(1.3) - dwellings considered as exclusions
(1.4) - dwellings located in commercial buildings
(1.5) - dwellings built after list established
(1.6) - forgotten or hidden dwellings

It should be noted that omitted seasonal dwellings were included with code (1.3). For a more detailed description of the codes, see the leaflet CLUSTER LISTING CHECK PROCEDURES appended at the end of this report.

Target Population and Description

The purpose of the program was to evaluate the listing of new clusters introduced in the LFS between October 1984 and March 1985, with the exception of clusters in the apartment and special area frames. The final total was 10,284 clusters; these constitute the population base on which the conclusions of this study were based.

For design and geographical reasons, the clusters were divided into three separate classes.

1. TRADITIONAL Clusters (T): are those found in NSRU urban areas or in small SRU towns, and correspond roughly to blocks, combinations of blocks or blockfaces. They are established in the same way as was done in the previous design.

2. BLOCKFACE Clusters (B): are those established by computer using the CENSUS BLOCKFACE FILE. They are located in large SRU cities, and like traditional clusters were generally blocks, combinations of blocks or blockfaces.

3. EA DESIGN Clusters (E): are clusters located outside the blockface areas of large cities and located in rural areas where the whole of an EA (or at least most of an EA) is the cluster as such (EA for enumeration area).

The breakdown into these three classes was carried out for each regional office (RO). The intersection of the classes with the ROs therefore allowed us to break the target population down into 24 strata.

Defining a Measure

In our search for a measure for describing a cluster, we felt that one point in particular was important. We believed that a measure is acceptable insofar as it assigns equal importance to both of the major types of possible errors. We therefore recommended the following measure.

N3 = number of dwellings erroneously included in the cluster listing

N1 = number of dwellings erroneously omitted from cluster listing

N0 = number of cluster dwellings prior to checking under this program

thus Y = (N0 - N3)/(N0 + N1) is a measurement of the number of cluster dwellings correctly identified by the first lister

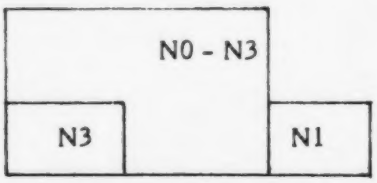
Example

After the cluster listing check, the following results were obtained:

- prior to check, 40 dwellings were listed
- after check, 5 dwellings erroneously included
  4 dwellings needed to be added

and Y = (40-5)/(40 + 4) = 0.79 (there were 9 changes in all).

In diagrammatic form, this gives



The measure taken will always be between 0 and 1; 0 if none of the dwellings listed belongs to the cluster, 1 if all of them are correctly listed. NO-N3 gives the number of dwellings correctly included on the list by the first lister. NO+N1 is then the maximum number of dwellings that could have been listed conditionally on N3, and the ratio of these two expressions defines our quality measure for a cluster.

Sampling

From the point of view of the accuracy of estimates, the size of the sample did not create any problem. Assuming that the values measured will vary between 0.5 and 1 (which is highly likely), it was relatively easy to achieve a 10% level of accurary for the estimates (see Appendix 1).

The problem was rather at the level of human resources, time and available budget.

Following a field mini-survey carried out by Head Office, it appeared that:

1. A total of 5 clusters would be the most effective workload to be covered by a checker working with BLOCKFACE clusters. This figure is based on criteria such as mileage in covering the clusters and time required to check them.

2. Similarly, a total of 4 clusters would be the optimum level for TRADITIONAL clusters.

3. Two clusters would be the optimum for the EA DESIGN class, on the condition that these 2 clusters were not too far apart.

4. The number of clusters to be checked should be as much as possible proportional to the size of the sample in each R.O.

Within these limitations, the proposed sampling plan was as follows:

1. BLOCKFACE
   - the number of SRUs appearing on the CENSUS BLOCKFACE FILE is 40
   - it would appear to be possible to sample each of the SRUs within the time and budgetary constraints involved
   - a sample of approximately 200, i.e., 5 randomly chosen clusters within each SRU, was drawn

2. TRADITIONAL
   - there are 248 small urban centres
   - it is estimated that approximately 50 urban centres could be sampled without going beyond the limits established. 48 is the figure that has been adopted
   - the size of the sample corresponding to each R.O. is then calculated (48 distributed proportionately to the number of urban centres by R.O.)
   - for each R.O., the number of clusters in an urban centre determines the size of the centre

- the centres (for each R.O.) are sampled using randomized systematic sampling with probability proportional to size
- finally, a random sample of 4 clusters is drawn from within each of these 48 centers
- a total of 192 clusters of this class were sampled

3. EA DESIGN

- there are 327 rural areas across Canada
- in view of the size of these areas (PSU), it is estimated that a maximum of 10 areas will be sampled per R.O.
- R.O.'s 12, 13, 15, 16, 17 will have this maximum of 10 areas sampled. Eight areas will be selected for R.O.'s 11 and 18
- for each area selected (the areas are selected in the same manner as the urban centres for the previous class), 2 clusters are selected at random
- the situation is different for R.O. 14. Because of the small number of areas, the clusters will be sampled directly (a single sampling level). It will be a simple random sample
- a total of 144 clusters were sampled

Data Collection

An initial check of the above mentioned 536 clusters was done by Head Office. At this stage the boundaries of each cluster were correctly identified. If an error was found, the cluster in question was excluded from the sample. Corrective action was then taken. This was the case for clusters 85106-68-001-5 and 63007-11-021-6.

For the other clusters sampled (534), a field check was carried out. This operation was carried out by the R.O.'s. Because of the time and budgetary constraints, cluster checking was limited to a check of the initial list (form F02). At this stage it is worth mentioning that it was preferable to assign the checking of a cluster to a person other than the one who made the initial list of dwellings in the

in the cluster. It should be remembered, of course, that the purpose of the program was in fact to assess the quality of the initial listing; a person who made an error in drawing up the list of dwellings is highly likely to make the same mistake when called upon to check his or her own list. A person with experience would be best suited to the task.

For the collection of the data, form F12, on which cluster updates are recorded, was modified. Column 16 was used to specify the error category in question (the various error categories are those mentioned earlier). The remaining columns remained unchanged. Column 16 is used only if a 1 or 3 is entered in column 11. Moreover the modified use of column 16 is valid insofar as it is used specifically for this program. Once the program is completed, column 16 will be used in the usual manner.

To summarize, checking a cluster, became an updating of the initial listing (F02) for the cluster, with all the usual procedures involved (i.e., the use of F12's right up to the data collection stage including form F57). Moreover if the figure entered in column 11 of form F12 is not 2, the checker must enter in column 16 the code corresponding to the error category.

With a view to eliminating any possible ambiguities in applying these new procedures, Head Office representatives visited each R.O. A procedures manual was distributed at that time (Appendix 4) and a number of clusters were checked during the visit. This made it possible for the checkers to familiarize themselves with the new program (note: in order to allow the Toronto R.O. visitor to check the EA DESIGN clusters, clusters 56005-24-000 and 56005-29-000 were added to the sample).

The final size of the sample was 535, because cluster 66030-27-000-6 was omitted.

Control of Errors

The quality of capture of data entered on form F12 was one of the key aspects in carrying out this program effectively. Unfortunately, it would have been extremely difficult for Head Office to carry out this task. As an experiment, we called upon the Montreal R.O. to return us a copy of the F12 forms used in the program. Comparing the data entered on the computer gave the following results:

RO	No Code (1)*	No Code (2)*	Omission	Change in Statust
13	26	11	1	6

Thus of 26 errors for which code 16, used for identifying error category, was not on the computer list, only 11 had been left out by the checker. The difference of 15 was attributable to errors in data entry. One error had been completely omitted and 6 had been entered incorrectly.

We sent the other R.O.'s a list of all errors involving the omission of code 16 from the computer list, to allow them to provide us with the missing data. Finally, we obtain the following table:

RO	No. Code (1)	No. Code (2)	Omission	Change	Total Number of Errors
11	17	2	2	0	61
12	30	12	0	0	60
13	26	11	1	6	205
14	4	-*	-*	-*	318
15	50	0	0	1	159
16	72	0	0	0	69
17	41	20	0	0	81
18	8	2	0	0	81

* Date entry for R.O. 14 was carried out at Headquarters.

** No Code (1): Number of errors for which code 16 was not on the computer list
No Code (2): Number of errors for which code 16 had not been left out by the checker

Before concluding this section, it should be pointed out that the 6 errors involving change in status at R.O. 13 had been coded with a 1 in space 16 of form F12. Moreover, these errors were introduced when the households in question entered the LFS. It is therefore likely that this coding error occurred on many occasions under similar circumstances.

Estimation

The estimation method varied somewhat depending on how the sample was drawn.

1. BLOCKFACE Clusters

Define the following notation:

- index i refers to the cluster, i = 1 to 5
- index j refers to the SRU, j = 1, ..., n_k
- index k refers to the R.O., k = 1, ..., 8
- n_k gives the number of SRUs in the R.O. k
- N_{kj} gives the number of clusters in SRU j of R.O. k
- W_{kj} gives the weighting of SRU j in R.O. k, $w_{kj} = N_{kj}/N_k$

then

$$\bar{y}_{kj.} = \sum_{i}^{5} \frac{y_{kji}}{5}$$

provides an estimate of the average listing quality for SRU j of R.O. k

$$v(\bar{y}_{kj.}) = (1 - F)S^2_{jk}/5$$

provides an estimate of the variance of $\bar{y}_{kj.}$ ($F = 5/N_{kj}$, $S^2_{kj} = \sum_{i} (y_{kji} - \bar{y}_{kj.})^2/4$)

$$\bar{y}_{k..} = \sum_{j}^{n_k} W_{kj}\bar{y}_{kj.}$$

provides an estimate of the listing quality for R.O. k

$$v(\bar{y}_{k..}) = \sum_{j} W^2_{kj}\, v(\bar{y}_{kj.})$$

provides an estimate of the variance of $\bar{y}_{k..}$

2. TRADITIONAL Clusters

Define the following:

- N_k is the number of cities in R.O. k
- n_k is the number of cities selected from R.O. k
- M_{kj} is the number of clusters in city j of R.O. k
- m_{kj} is the number of clusters selected from city j (here, $m_{kj} = 4 \; \forall \; k,j$)
- $s^2_{2kj} = \sum_i^4 (y_{kji} - \bar{Y}_{kj.})^2/3$
- π_{kj} is the probability that city j of R.O. k is in the sample
- π_{kji} is the probability that cities i and j of R.O. k are in the sample
- $\hat{Y}_{kj} = M_{kj} \sum_i y_{jki}/4$

then $\bar{y}_{kj.} = \dfrac{\hat{Y}_{kj}}{M_{kj}}$ provides an estimate of the average listing quality for city j of R.O. k

$\bar{y}_{k..} = \sum_j \dfrac{\bar{y}_{kj.}}{n_k}$ provides an estimate of the average listing quality for R.O. k

- if we let $\hat{Y}_{k.} = \sum_j \hat{Y}_{kj}/\pi_{kj}$

then

$$\hat{v}(Y_{k.}) = \frac{(1-\pi_{kj})}{\pi^2_{kj}} \hat{Y}^2_{kj} + 2 \sum_j \sum_{l<j} \hat{Y}_{kl} \hat{Y}_{kj} \frac{(1-\pi_{kj}\pi_{kl}/\pi_{kjl})}{\pi_{kj}\pi_{kl}} + \sum_j^{n_k} \frac{M_{kj}(M_{kj}-4)s^2_{2kj}}{4\pi_{kj}}$$

is an unbiased estimator of $V(\hat{Y}_{k.}) =$

$$\sum_{j=1}^{N_k} \frac{(1-\pi_{kj})Y^2_{jk}}{\pi_{kj}} + 2 \sum_j \sum_{l<j} \frac{(\pi_{kjl} - \pi_{kj}\,\pi_{kl})}{\pi_{kj}\pi_{kl}} Y_{kj} Y_{kl} + \sum_j^{N_k} \frac{M_{kj}(M_{kj} - 4)s^2_{2kj}}{4\pi_{kj}}$$

and $v(\bar{y}_{k..}) = v(\hat{Y}_{k.})/M^2_o$ $(M_o = \sum_j M_{kj},\ j = 1,\ N_k)$
is an estimate of the variance of $\bar{y}_{k..}$

However, since it is very difficult to calculate π_{kjl}, we use

$$v^*(\bar{y}_{k..}) = \sum_j (\bar{y}_{kj.} - \bar{y}_{k..})^2/n_k(n_k - 1),$$

which is an estimate of the variance of $\bar{y}_{k..}$ if the primary units are selected with replacement. We have $V^* > V$.

3. EA DESIGN Clusters

The estimates for these are the same as those above with $m_{kj} = 2\ \forall k, j$.

Each of the above expressions allows the direct calculation of the values of the average listing quality for the clusters for each possible stratum (R.O. class). To obtain estimates of this value at higher levels, (i.e., R.O., class, or Canada), the stratified sampling estimator may be used. For example if:

Z_i = estimate for stratum i

then

$$Z_{(Canada)} = \Sigma W_i Z_i \text{ and } v(Z_{(Canada)}) = \Sigma W^2_i\, v(Z_i),$$
where w_i = weight of stratum i in Canada.

Analysis

Table 1 shows the number of errors of each stratum (defined by the possible values for the R.O. (11, 12, 13, ..., 18) and class (B (Blockface), E (EA Design), T (Traditional)). Included are the number of clusters used in the study as well as the number of corrections made in this program.

Table 2 breaks down the number of corrections by error category. Column cterr gives the total number of corrections. The values appearing in columns N11 - N16, N31 - N34 correspond to the number of corrections for each possible

category. Each category is identified by comparing the values in columns 11 and 16 of the modified form F12. For example, N11 corresponds to the number of corrections resulting from the fact that a dwelling located outside the cluster was erroneously included (see appended manual). The total for the figures to the right of the column "cterr" is not equivalent to the value for the "cterr" column. This is because we were unable to obtain all the data required from the R.O.'s with respect to a number of clusters.

Table 3 gives the average quality as defined in section "DEFINING A MEASURE" of this study, of cluster listing for each possible division (see column YM). Under column YVV is an estimate of the square root of the variance for variable YM. This gives us an idea of the accuracy of the values obtained. Also, the remaining columns make it possible to pinpoint the source of the error. (EX: the first line of Table 3 shows that YM = 0.995, YVV = 0.005, YM16 = 0.005 and all the other variables are equal to zero. Since we have selected 5 clusters from this division (see Table 1), YM corresponds to the mean quality over these 5 clusters and

$$YM = (1 + 1 + 1 + 1 + 0.975)/5 = 0.995.$$

As we can see, 4 clusters out of 5 were found with no error. The last one had one error which is of the type N16 and YM16 = (0 + 0 + 0 + 0 + 0.025)/5 = 0.005.)

The average quality of clusters for Canada is estimated to be 98.5%. The quality for the various classes varies very little; approximately 1%. The same is true for the quality of the various R.O.'s. To identify whether these differences are due to something other than chance, a statistical test was applied to the data (the description of this test is given in Appendix 2). The results obtained did not allow us to distinguish the strata being studied, which suggests that the value estimated for Canada is an accurate reflection of the average quality of the clusters at all levels used in the study.

It is also interesting to note the value of YVV for 2 strata in particular. Stratum E14 and stratum T11 have values which appear to be high compared to the rest of the data. A glance shows that clusters 59112-65-010-4 and 02025-11-011-2 have measurements of quality equal to 49.2 and 63.2% respectively, which were the two lowest values recorded.

In view of the special nature of our data (i.e., the high percentage of errorless clusters), we decided to compare results using another criterion: another measure which takes the value 1 if no corrections are made to the cluster and a value of 0 otherwise. Table 4 gives the new estimates of the proportion of errorless clusters by stratum. The proportion of errorless clusters for Canada was estimated to be 72.8%. This figure varies considerably by class, ranging from 42.4% for class E to 80.2% for class B. The variations are smaller for R.O.'s. These estimates are also less accurate than the earlier measures.

Statistical tests were carried out with a view to analyzing these trends (see Appendix 3 for a description of the procedures followed). They show that class E is significantly different from the other 2 classes, and that the latter 2 are similar. There was no significant difference at the R.O. level (on the basis of available data). Nevertheless, a very low value was obtained for stratum E17.

It is interesting at this stage to contrast the data yielded by the 2 comparison methods described above.

The first comparison of mean quality of clusters by class indicated that the values obtained for classes E and T were similar. The second comparison, on the other hand, isolates only class E. It was therefore to be expected that the average quality for clusters containing at least one error should be lower for class T strata than class E strata. The following Table confirmed this expectation.

Estimates of quality of clusters with at least one error Table (A)

RO	11	12	13	14	15	16	17	18
BLOCKFACE	0,974	0,957	0,969	0,941	0,963	0,932	0,980	0,975
EA DESIGN	0,972	0,987	0,956	0,908	0,971	0,968	0,964	0,981
TRADITIONAL	0,808	0,895	0,949	0,985	0,949	0,911	0,916	0,950
ALL CLUSTERS	0,881	0,935	0,960	0,955	0,960	0,930	0,948	0,965

Estimates of mean number of errors percluster with at least one error Table (B)

RO	11	12	13	14	15	16	17	18
BLOCKFACE	1,000	2,074	1,859	4,360	4,252	22,078	1,500	2,452
EA DESIGN	2,917	2,417	11,944	36,571	9,111	2,071	3,450	3,500
TRADITIONAL	16,500	4,333	2,521	1,000	2,375	2,250	3,833	2,500
ALL CLUSTERS	10,110	3,306	3,610	6,170	4,352	10,137	2,855	2,582

As expected, the values for E14 and T11 are among the lowest in the Table (A). However, stratum E17 shows a high value, which means that despite the high level of incorrect clusters, the quality of defective clusters remained high. Because of the small amount of data available for each stratum, however, these trends should be taken as indicative only.

There may be a simple explanation for the fact that the mean quality of defective clusters for class T was lower than for other classes. Because class T clusters were generally smaller than the others, any error committed becomes relatively more significant in terms of the measure used (in the preceding Table (B), we see that the average number of errors per cluster is low for class T).

Conclusions and Recommendations

The quality of cluster listings for Canada was estimated to be 98.5% with a standard deviation of 0.2%. For the variables being studied, the mean quality of cluster listing does not vary (in a statistically significant fashion) by class of cluster (B, E, T) and R.O. (11, 12, ..., 18).

The proportion of errorless clusters varies by class. The proportion was estimated at 42.4% (4.8%) for class E, and 80.2% (3.4%) and 77.5% (2.9%) respectively for classes B and T. The result for class E is significantly different (alpha = 5%) from the other classes. The proportion of errorless clusters does not vary enough from one R.O. to another to make the identification of a trend possible.

The error categories and relations vary a great deal from one stratum to another. However, error (1, 6) (dwellings omitted in cluster enumeration because they were forgotten for no apparent reason or because they were difficult to find) would appear to be widespread in class E strata. A more detailed description of these relations is given in Appendix 5 of this report.

The quality of data entry could be improved. In fact, the value '1' for column 16 on the form F02 is confusing (when households added are included in the LFS, checkers tend to enter a '1' for column 16 regardless of the category of error encountered. The same is true for households to be dropped from LFS).

Although a great deal of effort has been expended to make appropriate use of the various error categories, it would appear that a number of checkers were uncertain about what to do in certain situations. A number of sub-codes were deliberately omitted because it was thought that their use was not appropriate. In any similar survey to be carried out in the future, these details will have to be emphasized still further.

More observations should be included to improve the power of the tests derived from variance analysis.

Acknowledgements

This paper has benefited from comments by and discussion with F. Mayda, J. Lindeyer, E. Reside, G. Keuhl and D. Drew.

CLUSTER LISTING CHECK PROGRAM

APPENDIX I

Determining the Sample Size

We shall assume that the sample drawn in a simple random sample. This makes it possible to obtain a good idea of the size required.

Thus
- X: variable under consideration
- $\bar{X}$ the mean for this variable in the population
- $\bar{x}$: the mean for this variable in a sample
- $cv(\bar{x})$: coefficient of variation for $\bar{x}$

We have $cv(\bar{x}) = \sqrt{V(\bar{x})/E^2(\bar{x})}$

$$V(\bar{x}) = \frac{(1-n/N \sum_{i=1}^{N} (X_i - \bar{X})^2}{(N-1)n}$$

$$\bar{X} = \sum^{N} X_i/N$$

Let us assume that we wish to have a level of accuracy such that $cv(\bar{x}) = 10\%$. We therefore have the following relation.

$$n = \frac{(1-n/N)S^2}{V(\bar{x})} = \frac{(1-n/N)S^2/\bar{X}^2}{V(\bar{x})/\bar{X}^2} = \frac{(1-n/N)S^2/\bar{X}^2}{(cv(\bar{x}))^2} = 100(1-n/N)S^2/\bar{X}^2$$

hence,

$$n = \frac{100S^2/\bar{X}^2}{(1+100S^2/\bar{X}^2N)}$$

The problem here boils down to the absence of data about the behaviour of the population in question ($S^2/\bar{X}^2$). Let us attempt to improve this measure. The basic assumption is that the variable measured has a value between a and b (in our case, a = 0.5 and b = 1).

Observation

Let G be a distribution of the values of the variable X, $X_i \in [a, b]$. Thus there exists a distribution G' with the same mean and a larger (or equal) variance such that X_i = a or b independently of i.

Suggested Proof

This involves showing that $\sum_{i=1}^{N} (X_i - \bar{X})^2$ is larger (or equal).
G'

To do so, let G be some distribution of X_i. Let us place these values in increasing order. Let X_{ib} be the largest value taken by variable X in the population, not equal to b. Let us make this value equal to b and subtract the difference $(b - X_{ib})$ from the lowest value for X not equal to a. This gives a distribution G' with the same mean and larger variance, by construction. The procedure is followed once again with G' as the initial distribution until G'n is reached, a distribution in which X_i = a or b independently of i.

Observation

Let there be population in which variable Y as measured takes as its value a or b, then the distribution that maximizes coefficient $S^2/\bar{Y}^2$ for the population is given by

$X = N\,(b/(b + a))$ X = number of observations with a as the value
N = population size

Proof

Maximize

$$\max \frac{\sum_i (Y_i - \bar{Y})^2}{N\bar{Y}^2} = \max \frac{\sum_i^N (Y_i - \bar{Y})^2}{\bar{Y}^2} = \max \frac{\sum_i^N Y_i^2 - N\bar{Y}^2}{\bar{Y}^2} = \max \frac{\sum_i^N Y_i^2}{\bar{Y}^2} - N =$$

$$\max \frac{\sum_i^N Y_i^2}{\bar{Y}^2} = \max \frac{\sum_i^N Y_i^2}{\frac{(\sum_i Y_i)^2}{N}} = \max F$$

But

$$\Sigma Y_i^2 = Xa^2 + (N - X)b^2 = X(a^2 + b^2) + Nb^2$$

$$\Sigma Y_i = Xa + (N - X)b = X(a - b) + Nb$$

$$(\Sigma Y_i)^2 = X^2 (a^2 - 2ab + b^2) + 2XNb (a - b) + N^2b^2$$

$$\frac{\partial F}{\partial X} = \frac{(X^2(a-b)^2+2XNb(a-b)+N^2b^2)\ (a^2-b^2)\ -\ (X(a^2-b^2+Nb^2)\,(2X(a-b)^2+2Nb(a-b)}{(X^2(a-b)^2+2XNB(a-b)+N^2b^2)}$$

= 0 if f (denominator not equal to 0 if at least one Y_i is not equal to 0)

$$X^2(a - b)^2 (a^2 - b^2) + 2XNb (a - b) (a^2 - b^2) + N^2b^2 (a^2 - b^2)$$

$$-2X^2(a - b)^2 (a^2 - b^2) -2XNb (a - b) (a^2 - b^2) - 2N^2b^2b (a - b) - 2XNb^2(a -b) = 0$$

$$-X^2(a - b)^2 (a^2 - b^2) - 2XNb^2 (a - b)^2 + b^2N^2 (a - b)^2 = 0$$

$$-X^2(a^2 - b^2) -2XNb^2 + b^2N^2 = 0 \qquad a \neq b$$

$$X = \frac{2Nb^2 \pm \sqrt{4N^2b^4+4(a^2-b^2)b^2N^2}}{-2(a^2-b^2)}$$

$$= \frac{2Nb^2 \pm 2Nab}{-2(a^2-b^2)} = \left[\frac{N\ (b^2 \pm ab)}{(a^2-b^2)}\right] = N\left[\frac{b}{a+b}\right]$$

because $b/(b-a) > 1$ is impossible.

Let us return to our formula for n. We have a distribution which maximizes $S^2/\bar{X}^2$: . 667N at 0.5 and .333N at 1.

$$S^2 = N/18\ (N - 1)$$

$\bar{X}^2 = 4/9$

and

$$\frac{S^2}{\bar{X}^2} = \frac{9N}{(N-1)72} = \frac{N}{(N-1)8} \cong \frac{1}{8}$$ (if N is large)

and

$n \cong 12.5/(1 + 12.5/N)$

Example

If N = 200 then n = 12 should allow us to obtain $cv(\bar{x}) = 10\%$.

If the case which concerns us, we want a $cv(\bar{x}) = 0.1$ for the R.O.s and classes, which is possible within our budget.

APPENDIX 2

Two Factor Variance Analysis

We attempt here to develop a test that will allow us to compare the different levels for the class and R.O. factors. There are 3 levels for class factor; 1 ('B'), 2 ('E'), 3 ('T').

There are 8 levels for the R.O. factor; 11, 12, 13, ..., 18.

One of the possible tests involves linear modelling of the observations measured. More precisely we are searching for an equation of the form

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \qquad \begin{cases} i = 1, 2, 3 \\ j = 1, 2, \ldots, 8 \\ k = 1, 2, \ldots, n \end{cases} \qquad (1)$$

where
- Y_{ijk} = value measured
- μ = overall mean
- t_i = effect of the i^{th} level of the class factor
- B_j = effect of thej^{th} level of R.O. factor
- $(\tau\beta)_{ij}$ = effect of interaction between T_i and B_j
- ϵ_{ijk} = error component

with the limitation

$$\sum_{i=1}^{a=3} \tau_i = 0, \quad \sum_{j=1}^{b=8} \beta_j = 0, \quad \sum_{i}^{a} (\tau\beta)_{ij} = \sum_{j}^{b} (\tau\beta)_{ij} = 0, \quad \epsilon_{ijk} \sim N(0, \sigma^2) \qquad (2)$$

The T_i, B_j therefore measure the deviation between the levels involved in the overall mean.

The index k denotes the number of repetitions per cell (i, j). Here, in class 1 ('B'), the sample measures is randomly divided into 2 independent sub-samples. This allows us to make 2 repetitions (the quality of the estimates remains excellent (see Table 5)).

The same is true for the other two classes, where each primary sampling unit is randomly assigned to a sub-sample. This, is therefore, an ANOVA with 2 observations per cell.

If we define

$$SST = \sum_{i}^{a}\sum_{j}^{b}\sum_{k}^{n=2} Y_{ijk}^2 - y^2.../abn$$

$$SSA = \sum_{i} Y_{i}^2../bn - y^2.../abn$$

$$SSB = \sum_{j} Y_{.j.}^2/an - y^2.../abn$$

$$SSAB = \sum_{i}\sum_{j} Y_{ij.}^2/n - y^2.../abn - SSA - SSB$$

$$SSE = SST - \left(\sum_{i}\sum_{j} Y_{ij.}^2/n - Y^2.../abn\right)$$

then

$$E(MSA) = E\left(\frac{SSA}{a-1}\right) = E\left(\frac{1}{a-1}\left[\sum_{}^{a}\frac{Y_{i..}^2}{bn} - \frac{Y_{...}^2}{abn}\right]\right)$$

$$= \frac{1}{a-1}E\left(bn\sum^{a}(\bar{Y}_{i..} - \bar{Y}_{...})^2\right) \text{ by substituting in (1) and (2)}$$

$$= \frac{bn}{a-1}E\left(\sum_{i=1}^{a}(\mu + \tau_i + \bar{\varepsilon}_{i..} - \mu - \bar{\varepsilon}_{...})^2\right)$$

$$= \frac{bn}{a-1}E\left(\sum_{i=1}^{a}\tau_i^2 - 2\sum_{i=1}^{a}\tau_i\bar{\varepsilon}_{...} + 2\sum_{i=1}^{a}\tau_i\bar{\varepsilon}_{i..} + \sum_{i=1}^{a}\bar{\varepsilon}_{i..}^2 - 2\sum_{i=1}^{a}\bar{\varepsilon}_{i..}\bar{\varepsilon}_{...} + \sum_{i=1}^{a}\bar{\varepsilon}_{...}^2\right)$$

$$= \frac{bn}{a-1}E\left(\sum_{i=1}^{a}\tau_i^2 - 0 + 2\sum_{i=1}^{a}\tau_i\bar{\varepsilon}_{i..} + \sum_{i=1}^{a}\bar{\varepsilon}_{i..}^2 - 2a\bar{\varepsilon}_{...}^2 + a\bar{\varepsilon}_{...}^2\right)$$

But $\varepsilon_{ijk} \sim N(0, \sigma^2) \Rightarrow E(\bar{\varepsilon}_{...}^2) = \sigma^2/abn \qquad E(\bar{\varepsilon}_{i...}^2) = \sigma^2/bn$

$$= \frac{bn}{a-1}\left(\sum_{i=1}^{a}\tau_i^2 + \frac{a\sigma^2}{bn} - \frac{a\sigma^2}{abn}\right)$$

$$= \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

and similarly, we find

$$E(MSB) = \sigma^2 + \frac{an}{b-1} \Sigma \beta_j^2$$

$$E(MSAB) = \sigma^2 + \frac{n\Sigma\Sigma(\tau_\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MSE) = \sigma^2$$

By relating the latter quantities to E (MSE), we obtain tests that allow us to compare the levels for the class and R.O. factors.

After checking with the Kolomogorov-Smirnov d - statistic that the ε_{ijk} have a normal distribution with mean 0 and variances σ^2 we obtain the following results

Dependent Variable : Estimate of Cluster Listing Quality

Source	d.f.	Sum of Squares	Weighted Sum of Squares	F	pr > F
model	23	0.00735167	0.00031964	0.93	0.572
error	24	0.00828500	0.00034521	R^2	
total corr.	47	0.01563667		0.47	

Source	d.f.	Sum of Squares	Weighted Sum of Squares	F	pr > F
class	2	0.00103517		1.5	0.2435
R.O.	7	0.00168800		0.7	0.6729
class R.O.	14	0.00462850		1.0	0.5189

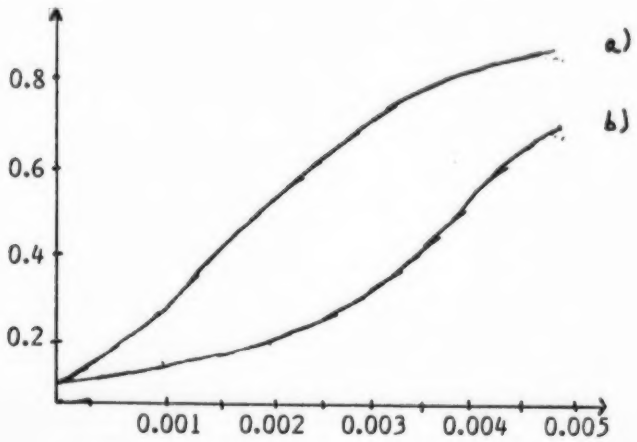
None of the effects is significant at the alpha = 5% level. The model therefore depends on a single parameter, the overall mean ($y = \mu + \epsilon$). This means that the variables used (class, R.O.) do not explain the behaviour of the values measured (quality of cluster listing) at the alpha = 5% level.

Power of the Test

a) H_o : $\tau_i = 0$ $i = 1, 3$ b) H_o : $\beta_j = 0$ $j = 1, \ldots, 8$

H_1 : $\tau_i \neq 0$ for at least one i H_1 : $\beta_j \neq 0$ for at least one j

We will restrict ourselves to giving one power curve for each of the 2 above situations.



abscissa: hypothetical sum of squares under H1
ordinate: probability of rejecting Ho given H1

For scale, let us add that

if SSA = SSB = 0,001 then the following relation $\sqrt{\frac{\Sigma \tau_i^2}{a\sigma^2}} = \sqrt{\frac{\Sigma \beta_j^2}{b\sigma^2}} = 1/4$

if SSA = SSB = 0,004 then the following relation $\sqrt{\frac{\Sigma \tau_i^2}{a\sigma^2}} = \sqrt{\frac{\Sigma \beta_j^2}{b\sigma^2}} = 1/2$

(Recall that $\hat{\tau}_i = (\bar{y}_{i..} - \bar{y}_{...})$, $\hat{\beta}_j = (\bar{y}_{.j.} - \bar{y}_{...})$)

APPENDIX 3

Two Factor Variance Analysis

As in Appendix 2, we attempt here to describe the observations using a linear statistical model. Howerver, it is impossible in this instance to consider 2 independent sub-samples without sacrificing too much accurary in the estimates. The proposed model is as follows

$$Y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij} \qquad \begin{cases} i = 1, \ldots, 3 = a \\ j = 1, \ldots, 8 = b \end{cases}$$

Hence, this is an ANOVA with one observation per cell. The expected values for the weighted sum of squares E (MSA), E (MSB), E (MSAB) are the same as for the preceding case with n = 1. The problem here is that it is impossible to obtain an estimate of E (MSE) (because there are no more degrees of freedom remaining). The experimental error and the interaction term become confused and it is impossible to test the factors unless the interaction term is 0. If this is the case, the model then becomes

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases}$$

and we estimate σ^2 by E (MSAB).

To test whether the interaction term is 0, Tuckey (1949) developed the following test

$$\text{if } SSN = \left[\frac{\sum\sum Y_{ij} Y_{i.} Y_{j.} - Y_{..} (SSA + SSB + Y_{..}^2 / ab)}{ab\,SSA\,SSB} \right]^2$$

then $SSN \Big/ (SSAB)/((a-1)(b-1))$ has a Fisher 1, (a-1)(b-1) - 1 distribution (under H_o : $(\tau\beta)_{ij} = 0 \ \forall_{ij}$)

and we obtain the value 1,00555 and $P(Fish_{(1,\ 13)} > 1{,}00555) > 25\%$. We do not reject H_o.

After using the Kolomogorov-Smirnov d - statistic to check that the ϵ_{ij}'s have a normal distribution with mean 0 and variance sigma squared, we obtain the following results.

Dependent Variable : Estimates of the Proportion of Errorless Clusters

Source	d.f.	Sum of Squares	Weighted Sum of Squares	F	pr > F
model	9	0.84135258	0.09348362	5.22	0.003
error	14	0.25062475	0.01790177	R^2	
total corr.	23	1.09197733		0.77	

Source	d.f.	Sum of Squares	Weighted Sum of Squares	F	pr > F
class	2	0.77025058		21.5	0.0001
R.O.	7	0.07110200		0.6	0.7706

Here, the effect of the class factor is significant. The model therefore reduces to $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ and explains approximately 75% of the total variability of data. The Duncan rank test is then applied to the class factor values. It shows that class E is significantly different (alpha = 5%) from the other classes, and that the other two classes are similar.

The tests used in appendix II and III to establish the above results assume that each stratum (i.e., the intersection of a class with an R.O.) has the same weight in terms of analysis.

Appendix 4

CLUSTER LISTING

CHECK

PROCEDURES

1. Purpose

The purpose of this program is to measure and evaluate the overall quality of field listing for the 1981 Redesign implementation. Many activities are associated with the task of listing these dwellings such as, obtaining area maps, field counting, preparing F01 diagrams, data capture, and so on, however, your task will be concentrated on the actual listing portion.

The Labour Force Survey is a sample survey, sometimes referred to as a representative sample which means that a dwelling selected for the survey could represent up to 120 or more typical dwellings in the area. For this reason, it is important that every dwelling within the boundaries of selected areas and only those within the boundaries be included at the time of listing.

2. Objectives

The objective of the program is not only to add or delete dwellings which should or should not have been listed, but through analysing noted discrepancies to be able to improve established procedures or develop new ones.

3. Selecting Clusters

Clusters selected for the program will be made up of three types of clusters; traditional clusters, which are those found in small S.R.U. areas and urbans areas, blockface clusters, which were primarily formed by computer in large S.R.U.'s and E.A. Design clusters, which are rural areas where whole or large parts of enumeration area's have been used as clusters. Due to their unique characteristics, special area and apartment clusters have been excluded.

4. General

Each regional office will decide who will perform the listing checks depending on travel arrangements, distance, cost and so on, however the

following conditions must be followed. The listing checker must be:

- someone other than the original lister
- experienced in listing and/or
- familiar with listing procedures
- either LF personnel, Senior LF Interviewer or LF Interviewer

5. **Timing**

Cluster checks will be conducted from early October through to mid December. You will be advised by your regional office on the amount of time (allocated) to complete each individual cluster.

6. **Preparation**

Before you leave your home/office you should:

- make sure you have proper area maps
- study and plan your route
- record your time and mileage
- make sure the F01 indicates plotted dwellings (if applicable)
- bring along a supply of Form 12's

Upon arriving at the cluster location (it is important that you) verify your starting point, by identifying at least 2 landmarks, for example, railway tracks, street name, hydro-lines, township markers, etc., then proceed in the same fashion as you would if you were actually listing the cluster.

7. **Verification**

When you receive your assignment, verify that the control data has been properly entered on a Record of Cluster Check Form 57. Should there be a discrepancy contact your regional office immediately.

a) No Differences

If, no additions or deletions are found during the cluster check, other than minor changes, initial and date the Record of Cluster Check Form 57.

b) Differences Noted

If, while conducting a cluster check, you discover a dwelling that was not listed and should have been or, visa versa, a dwelling that was listed and should not have been, complete a Cluster List Update Form 12 in the normal fashion along with a special note that is to be entered in Item 16 of the Form 12 to best describe the reason for the disparity.

NOTE: For the purpose of this program we are not interested in so called multiples. However, should you discover what might be a "multiple dwelling", simply attach a note to the Cluster list to alert the regional office.

8. Special Codes

Below you will find a list of codes, along with a brief explanation. One of these codes should be entered in Item 16 of the Form 12 anytime a dwelling is added or deleted to the cluster list.

Newly Discovered Dwelling

Code 1: Cluster Boundary Misinterpretation (N11)

Use Code 1 for dwellings that were missed because the Lister did not properly identify the boundaries.

Code 2: Portion of Road Network Missed (N12)

Use Code 2 for missed dwellings that are located well within the boundary limits and cannot be classified as Code 1 (Boundary Misinterpretation). This can usually be determined when one or more consecutive dwellings were missed.

Code 3: Dwellings Thought to be Exclusion (N13)

Use Code 3, if it appears that the missed dwellings were presumed to be exclusions by the Lister for example; apartment sample special area or an area which was identified on the F01 diagram as being an "isolated cottage area".

Code 4: Dwellings Located in Business Structures (N14)

Use Code 4 when it appears that a dwelling(s) was missed due to "business related" reasons for example; a dwelling attached to a business, apartments in an office building.

Code 5: New Dwellings Since Original Listing (N15)

Use Code 5, if it appears that a new dwelling has been constructed since the time of listing. This also applies to a dwelling which has been converted from a business to a private residence since the time of listing.

Code 6: Missed or Hidden Dwelling (N16)

Code 6 applies to dwellings that were missed for no apparent reason or were hidden from view. Before using this code, it must be evident from the listing that this portion of the road was covered by the Lister.

Deleted Dwellings

Code 1: Boundary Misinterpretation (N31)

This code is to be used, when it has been positively established that a dwelling which was listed, is actually located outside the cluster.

Code 2: Dwelling Converted to Business (N32)

Use Code 2 for a dwelling which is used solely for business purposes only and has no perminent residents.

Code 3: Dwelling Demolished or Burned Down (N33)

Code 3 applies to any dwelling, which was originally listed, that has either been demolished, burned down or relocated since the time of listing.

Code 4: Dwelling Incorrectly Included (N34)

Enter Code 4 if:

- a dwelling is listed more than once
- an apartment building contains less dwelling than listed
- a collective dwelling in a Special Area Frame was listed
- the dwellings listed are in a building belonging to the apartment sample

9. Returning your Assignment

Completed assignment must be returned to the regional office in one shipment and should include all the material originally received, a completed Project Claim (F85) and the Record of Cluster Check (F57) indicating the completion of the checks.

10. Points to Remember

All clusters, no matter how large or small, are subject to change even over short periods of time. However, remembering the following points, will ensure a high quality of listing. While performing Cluster Checks you should:

- identify cluster boundaries at all times
- proceed on foot in high density areas
- always verify isolated cottage areas do not contain any permanent residences
- complete a Cluster List Update (Form 12) for additions or deletions only.

Appendix 5

Description of errors

The question that concerns us here is to know whether one category of errors occurs most often when a correction is made to the cluster.

To answer this question, a more careful examination of the behaviour of each stratum is required because of both the low number of errors in certain cells and interaction.

But first it should be pointed out that there are 2 ways to approach this question: (1) an attempt can be made to answer the question on the basis of the number of dwellings incorrectly included or excluded by error category (which is shown in Table 2), (2) another method would be to do an analysis of the frequency of the various categories of errors for each cluster in which errors are found.

To illustrate these 2 approaches let us examine a fictitious example in which, of the 30 clusters being studied, 13 include at least one error. Let us assume that Table 2, for this example of 30 clusters, is as followed:

CTERR	N11	N12	N13	N14	N16	N31	N32	N34
48	0	3	2	4	13	0	2	24

In addition, let us assume that the 13 category (1, 6) errors (which correspond to N16) are distributed in 10 of 13 clusters found to be incorrect, while all remaining errors come from a single cluster.

This example shows clearly that it is important to consider both of the above described aspects. In fact, although half the errors were identified in category (3, 4), only one of 13 clusters is affected by this type of error, compared to 10 of 13 by category (1, 6), which suggests a more widespread problem for the latter. We would then have to conclude that of the 15 categories of corrections used by the interviewer (once to describe the errors under N12, once to describe N13, one to describe N14, 10 times to describe N16, once to describe N32, once to describe N34), 10 are of the (1, 6) category.

In this analysis which follows, each of the two aspects is therefore included. There is little to be said about strata B11, B12, T14, T15, T16 and T17 because of the small number of observations recorded.

Stratum B13 - Slightly more than one-third of errors are of category (3, 4). Moreover, 5 of the 9 clusters found to be incorrect included this category of error. Special attention must be given to this category.

Stratum B14 - Here, there are 3 categories of errors to be considered: (1, 2), (1, 6) and (3, 4). Of the 21 categories of corrections used, 19 belong to one of these categories, i.e., 5, 8 and 6, respectively. Special attention is called for here.

Stratum B15 - 37 of the 56 errors recorded belong to category (1, 3). However, a study of the data shows that all these errors stem from a single cluster; 52103-61-0124. The remaining errors are equally distributed. The explanation for the dubious cluster is referred to the R.O.

Stratum B16 - Same behaviour as previous stratum. The defective cluster is 67154-63-034. It should be added that only 2 clusters were incorrect and there is not much to be said about this.

Stratum B17 - No notable trends.

Stratum B18 - 5 of 7 categories of corrections used are of category (1, 6). The 5 N12 errors stem from a single cluster; 93101-65-019. Special attention must therefore be given to category (1, 6).

Stratum E11 - 15 of 25 errors are of category (1, 6). Also, half (6/12) of the categories of correction used are of this category. Special attention is required.

Stratum E12 - As for the previous stratum, 6 of the 9 categories used are of category (1, 6). The same recommendations apply.

Stratum E13 - The value 74 in cell (1, 3) is largely the result of cluster 43044-22-000. To be specific, the error is due to the fact that the interviewer did not enter seasonal dwellings on list F02. It would perhaps be a good idea for this R.O. to clarify the situation with its interviewers. However, the (1, 6) error category is much more widespread for various clusters (10 of 28 categories used). Special attention should be paid. Because the size of the clusters in this class is large, the value of 18 is not significant.

Stratum E14 - 231 of the 234 errors in cell (3, 1) result from cluster 59112-65-010. Categories (1, 6) and (3, 1) should be given careful attention. Of the 11 categories used, 8 are related to these 2 categories, i.e., 5 and 3 respectively.

Stratum E15 - The 33 errors in cell (1, 1) result from cluster 53681-64-041. However, categories (1, 6) and (3, 4) with 15 (10 and 5) of the 23 categories used are those that require the most attention.

Stratum E16 - No specific trends noticed.

Stratum E17 - Once again, category (1, 6), with 11 of the 18 correction categories used, is the major problem. Also, category (1, 1) is worthy of attention.

Stratum E18 - Category (1, 6), with 11 of the 19 correction categories used, requires attention.

To summarize, for class E, category (1, 6) is clearly the most frequent one.

Stratum T11 - Cell (3, 1) results from a single cluster; 2035-11-011.

Stratum T12 - The 9 errors for cell (1, 1) result from cluster 22101-61-017. The remaining error categories use are distributed evenly.

Stratum T13 - 7 of 18 categories used relates to cell (3, 4). Also, 5 other categories relate to cell (1, 6) with 15 of the 48 errors attributable to cluster 49201-63-075. Special attention should be given to these 2 categories of errors.

Stratum T18 - No specific trends noted.

TABLE 1

OVERVIEW OF LISTING CHECK PROGRAM

		LISTLAST	C13	DSN	LISTLAST	CTERR
		NUMBER OF CLUSTERS IN TEST	No. OF CLU. WITH ADDITIONS OR DELETIONS	TOTAL DESIGN COUNT	TOTAL LINES ORINALLY LISTED	TOTAL LINES ADDED OR DELETED
Class	R.O.					
Blockface	St. John	5	1	449	447	1
	Halifax	15	3	874	864	7
	Quebec	35	9	2469	2678	20
	St. Falls	35	16	2156	2578	61
	Toronto	60	11	3594	3844	56
	Winnipeg	10	2	887	1013	36
	Edmonton	15	5	944	1023	11
	Vancouver	25	6	1759	1947	17
EA Design	St. John	16	8	1601	1943	27
	Halifax	20	8	2247	2707	22
	Quebec	20	14	4736	5262	137
	St. Falls	12	7	1563	2024	256
	Toronto	22	12	4096	4929	88
	Winnipeg	19	10	1267	1534	25
	Edmonton	20	16	2138	2706	61
	Vancouver	16	11	2638	3083	46
Traditional	St. John	16	2	650	694	33
	Halifax	24	8	1095	1126	31
	Quebec	40	13	1900	1926	48
	St. Falls	12	1	737	734	1
	Toronto	24	5	1018	1024	15
	Winnipeg	23	3	699	695	8
	Edmonton	27	4	1035	1016	14
	Vancouver	24	8	1003	1060	23
Blockface	Canada	200	53	13132	14394	209
E.A. Design	Canada	145	86	20286	24188	662
Traditional	Canada	190	44	8137	8275	173

		LISTLAST	C13	DSN	LISTLAST	CTERR
		NUMBER OF CLUSTERS IN TEST	No. OF CLU. WITH ADDITIONS OR DELETIONS	TOTAL DESIGN COUNT	TOTAL LINES ORINALLY LISTED	TOTAL LINES ADDED OR DELETED
Class	R.O.					
Canada	St. John	37	11	2700	3084	61
	Halifax	59	19	4216	4697	60
	Quebec	95	36	9105	9866	205
	St. Falls	59	24	4456	5336	318
	Toronto	106	28	8708	9797	159
	Winnipeg	52	15	2853	3242	69
	Edmonton	62	25	4117	4745	86
	Vancouver	65	25	5400	6090	86
Canada	Canada	535	183	41555	46857	1044

TABLE II

NUMBER OF ERRORS PER CATEGORIES

		CTERR SUM	N11 SUM	N12 SUM	N13 SUM	N14 SUM	N16 SUM	N31 SUM	N32 SUM	N34 SUM
CLASS	R.O.									
Blockface	St. John	1	0	0	0	0	1	0	0	0
	Halifax	7	0	0	0	0	1	0	1	5
	Quebec	20	0	0	0	2	0	3	1	7
	St. Falls	61	0	21	0	0	19	3	1	17
	Toronto	56	1	3	37	0	4	2	1	7
	Winnipeg	36	0	0	34	0	0	0	0	2
	Edmonton	11	0	0	0	1	2	2	2	4
	Vancouver	17	0	5	0	0	10	0	0	1
E.A. Design	St. John	27	1	5	1	2	15	0	1	0
	Halifax	22	0	2	0	0	10	1	1	0
	Quebec	137	4	18	74	2	30	2	3	4
	St. Falls	256	1	0	0	2	13	234	0	2
	Toronto	88	33	22	0	0	16	6	3	8
	Winnipeg	25	2	2	3	1	5	1	3	8
	Edmonton	61	19	0	0	0	28	2	2	1
	Vancouver	46	1	9	1	1	28	4	1	1
Traditional	St. John	33	0	0	0	1	0	32	0	0
	Halifax	31	9	0	0	0	5	5	0	9
	Quebec	48	0	4	0	5	25	1	2	7
	St. Falls	1	0	0	0	0	1	0	0	0
	Toronto	15	0	0	0	0	7	0	0	8
	Winnipeg	8	0	0	0	3	1	0	3	1
	Edmonton	14	0	0	0	0	1	0	0	0
	Vancouver	23	0	8	6	3	2	0	3	0
Canada		1044	71	99	156	23	224	298	28	92

Table III

MEANS BY TYPE OF CLUSTER DESIGN

		YM SUM	YVV SUM	YM11 SUM	YM12 SUM	YM13 SUM	YM14 SUM	YM16 SUM	YM31 SUM	YM32 SUM	YM34 SUM
CLASS	R.O.										
Blockface	St. John	0.995	0.005	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000
	Halifax	0.991	0.006	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.007
	Quebec	0.995	0.001	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.002
	St. Falls	0.965	0.012	0.000	0.020	0.000	0.000	0.004	0.001	0.000	0.010
	Toronto	0.997	0.001	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001
	Winnipeg	0.986	0.013	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.001
	Edmonton	0.990	0.005	0.000	0.000	0.000	0.001	0.001	0.002	0.002	0.003
	Vancouver	0.996	0.002	0.000	0.001	0.000	0.000	0.002	0.000	0.000	0.000
	Canada	0.991	0.002	0.000	0.002	0.002	0.000	0.001	0.001	0.001	0.003
E.A. Design	St. John	0.986	0.006	0.001	0.003	0.001	0.001	0.007	0.000	0.000	0.000
	Halifax	0.994	0.003	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
	Quebec	0.974	0.012	0.001	0.003	0.014	0.000	0.007	0.000	0.001	0.001
	St. Falls	0.946	0.037	0.000	0.000	0.000	0.001	0.005	0.044	0.000	0.001
	Toronto	0.986	0.006	0.005	0.003	0.000	0.000	0.003	0.001	0.001	0.001
	Winnipeg	0.979	0.011	0.001	0.001	0.001	0.000	0.007	0.002	0.003	0.006
	Edmonton	0.968	0.011	0.006	0.000	0.000	0.000	0.017	0.001	0.001	0.000
	Vancouver	0.985	0.006	0.000	0.002	0.000	0.000	0.010	0.001	0.000	0.000
	Canada	0.981	0.003	0.002	0.001	0.002	0.000	0.007	0.002	0.001	0.001
Traditional	St. John	0.976	0.023	0.000	0.000	0.000	0.001	0.000	0.023	0.000	0.000
	Halifax	0.969	0.010	0.010	0.000	0.000	0.000	0.005	0.004	0.000	0.008
	Quebec	0.978	0.009	0.000	0.003	0.000	0.003	0.008	0.000	0.001	0.004
	St. Falls	0.999	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
	Toronto	0.986	0.010	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.007
	Winnipeg	0.987	0.012	0.000	0.000	0.000	0.005	0.002	0.000	0.005	0.002
	Edmonton	0.986	0.008	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
	Vancouver	0.981	0.009	0.000	0.006	0.006	0.002	0.003	0.000	0.002	0.000
	Canada	0.981	0.004	0.002	0.001	0.001	0.001	0.003	0.002	0.001	0.003

MEANS BY TYPE OF CLUSTER DESIGN

CLASS	R.O.	YM SUM	YVV SUM	YM11 SUM	YM12 SUM	YM13 SUM	YM14 SUM	YM16 SUM	YM31 SUM	YM32 SUM	YM34 SUM
Canada	St. John	0.982	0.013	0.000	0.001	0.000	0.001	0.003	0.013	0.000	0.000
	Halifax	0.981	0.005	0.005	0.000	0.000	0.000	0.004	0.002	0.000	0.006
	Quebec	0.986	0.004	0.000	0.002	0.002	0.002	0.004	0.000	0.001	0.003
	St. Falls	0.976	0.007	0.000	0.010	0.000	0.000	0.003	0.005	0.000	0.006
	Toronto	0.993	0.003	0.001	0.001	0.001	0.000	0.002	0.000	0.000	0.002
	Winnipeg	0.985	0.007	0.000	0.000	0.005	0.002	0.002	0.000	0.003	0.002
	Edmonton	0.085	0.005	0.001	0.000	0.000	0.000	0.003	0.001	0.001	0.001
	Vancouver	0.988	0.004	0.000	0.003	0.003	0.001	0.003	0.000	0.001	0.000
	Canada	0.985	0.002	0.001	0.001	0.002	0.001	0.003	0.002	0.001	0.002

TABLE IV PROPORTIONS BY TYPE OF CLUSTER DESIGN

		YM SUM	YVV SUM
Classe	R.O.		
Blockface	St. John	0.800	0.194
	Halifax	0.800	0.118
	Québec	0.857	0.034
	St. Falls	0.481	0.089
	Toronto	0.908	0.023
	Winnipeg	0.800	0.145
	Edmonton	0.700	0.121
	Vancouver	0.890	0.047
	Canada	0.802	0.034
E.A. Design	St. John	0.500	0.134
	Halifax	0.600	0.125
	Québec	0.300	0.111
	St. Falls	0.417	0.132
	Toronto	0.455	0.106
	Winnipeg	0.450	0.139
	Edmonton	0.200	0.082
	Vancouver	0.313	0.132
	Canada	0.424	0.048
Traditional	St. John	0.875	0.072
	Halifax	0.667	0.053
	Québec	0.675	0.084
	St. Falls	0.917	0.083
	Toronto	0.792	0.077
	Winnipeg	0.875	0.085
	Edmonton	0.845	0.078
	Vancouver	0.667	0.083
	Canada	0.775	0.029
Canada	St. John	0.760	0.064
	Halifax	0.675	0.051
	Québec	0.706	0.039
	St. Falls	0.640	0.058
	Toronto	0.826	0.027
	Winnipeg	0.764	0.073
	Edmonton	0.703	0.062
	Vancouver	0.735	0.044
	Canada	0.728	0.020

TABLE V MEANS BY TYPE OF CLUSTER DESIGN

		YM SUM	YVV SUM	YM SUM	YVV SUM
Classe	R.O.				
Blockface	St. John	1.000	0.000	0.987	0.013
	Halifax	1.000	0.000	0.985	0.010
	Québec	0.996	0.001	0.994	0.001
	St. Falls	0.938	0.020	0.982	0.005
	Toronto	0.996	0.002	0.998	0.001
	Winnipeg	1.000	0.000	0.967	0.032
	Edmonton	0.990	0.009	0.991	0.002
	Vancouver	0.999	0.001	0.993	0.004
	Canada	0.992	0.002	0.989	0.005
E.A. Design	St. John	0.989	0.002	0.982	0.013
	Halifax	0.990	0.005	0.997	0.001
	Québec	0.962	0.023	0.987	0.008
	St. Falls	0.987	0.007	0.905	0.073
	Toronto	0.991	0.005	0.981	0.012
	Winnipeg	0.996	0.002	0.963	0.020
	Edmonton	0.980	0.009	0.956	0.019
	Vancouver	0.989	0.007	0.981	0.010
	Canada	0.985	0.004	0.976	0.006
Traditional	St. John	0.954	0.046	0.998	0.002
	Halifax	0.980	0.009	0.958	0.016
	Québec	0.982	0.012	0.974	0.016
	St. Falls	0.998	0.002	1.000	
	Toronto	0.978	0.020	0.994	0.003
	Winnipeg	0.997	0.003	0.976	0.024
	Edmonton	0.976	0.012	1.000	0.000
	Vancouver	0.978	0.018	0.983	0.010
	Canada	0.981	0.005	0.983	0.005
Canada	St. John	0.971	0.025	0.992	0.004
	Halifax	0.987	0.005	0.975	0.008
	Québec	0.986	0.006	0.986	0.006
	St. Falls	0.965	0.011	0.981	0.008
	Toronto	0.991	0.005	0.995	0.002
	Winnipeg	0.998	0.001	0.970	0.017
	Edmonton	0.982	0.007	0.991	0.003
	Vancouver	0.989	0.008	0.988	0.005
	Canada	0.986	0.003	0.984	0.003